

# **Evaluating Rolling Dose Designs and Methods**

**Amit Roy and Frank Shen**

Bristol-Myers Squibb

Joint Statistical Meeting, Seattle, WA

August 6-10, 2006

## Acknowledgments: Rolling Dose Study Working Group

---

- Alex Dmitrienko, Eli Lilly
- Amit Roy, BMS
- Beat Neuenschwander, Novartis
- Björn Bornkamp, U. of Dortmund
- Brenda Gaydos, Eli Lilly
- Chyi-Hung Hsu, Pfizer
- Frank Bretz, Novartis
- Frank Shen, BMS
- Franz König, Med. U. Vienna
- Greg Enas, Eli Lilly
- José Pinheiro, Novartis <sup>a</sup>
- Michael Krams, Pfizer
- Qing Liu, J & J
- Rick Sax, AstraZeneca <sup>a</sup>
- Tom Parke, Tessella

---

<sup>a</sup>Leaders

- Background and motivation
- Rolling Dose Studies PhRMA initiative: goals and scope
- Evaluation of designs and methods for dose finding:  
simulation study
- Simulation results
- Discussion

## Background

---

- Pharmaceutical industry **pipeline problem**: decreasing number of drug approvals, despite advances in basic science
- FDA's **Critical Path Initiative**: “Innovation vs. Stagnation” White Paper
- Pharmaceutical industry (PhRMA) response: working groups (WGs) to address **key drivers** of poor performance
- Rolling Dose Studies (RDS) WG formed to stimulate innovation in **dose finding** studies

## Motivation: RDS for dose finding

---

- Poor understanding of dose response (DR) for both **efficacy** and **safety** is pervasive in drug development
- Sub-optimal dose selection identified by both FDA and industry as one of **root causes** of late phase attrition and post-marketing problems with approved drugs
- Current dose finding designs and methods focus on selection of registrational study dose out of fixed, generally small number of dose levels, via pairwise hypothesis testing  $\implies$  **inefficient**

## Goals of RDS initiative

---

- Investigate and develop designs and methods for efficiently **learning** about safety and efficacy dose-response  $\implies$  benefit/risk profile
- More accurate and faster **decision making** on dose selection and improved labeling
- Evaluate statistical operational characteristics of alternative designs and methods to make recommendations on their use in practice
- Increase awareness about this class of designs, promoting their use, when advantageous

## Definition and Scope of RDS

---

- Flexible dose-ranging designs allowing **dynamic** allocation of patients and possibly variable number of dose levels based on accumulating information
- Intended to strike **balance** between need for additional DR information and increased costs and time-lines
- Emphasis on modeling/estimation (**learning**) as opposed to hypothesis testing (**confirming**)
- Investigate existing and new RDS methods via simulation
- Evaluate potential benefits over traditional dose-ranging designs over variety of scenarios to make recommendations on practical usefulness of RDS methods

## Dose Finding Methods – Fixed Doses

---

- Traditional **ANOVA** based on pairwise comparisons and multiplicity adjustment (Dunnett); common approach used in dose finding studies – Amit Roy and Frank Shen
- **MCP-Mod** combination of multiple comparison procedure (MCP) to identify presence of DR and modeling, to estimate target dose(s) and DR profile (Bretz, Pinheiro and Branson, 2005) – José Pinheiro and Frank Bretz
- **MTT**: novel method based on Multiple Trend Tests – developed by Qing Liu
- Bayesian Model Averaging: **BMA** – Beat Neuenschwander and Amy Racine
- Nonparametric local regression fitting: **LOCFIT** – Björn Bornkamp and Frank Bretz

## Dose Finding Methods – RDS

---

- **GADA**: Dynamic dose allocation based on Bayesian normal dynamic linear model (Krams, Lees and Berry, 2005); allocation of patients to dose adaptively changed according to model-based optimization criteria (e.g., variance of target dose estimate) – Tom Parke and Michael Krams
- **D-opt**: adaptive dose allocation based on D-optimality criterion used with sigmoid- $E_{\max}$  model; model parameters re-estimated at interim analysis and corresponding D-optimal allocation determined for next interval – developed by Alex Dmitrienko; Chyi-Hung Hsu helped with simulations

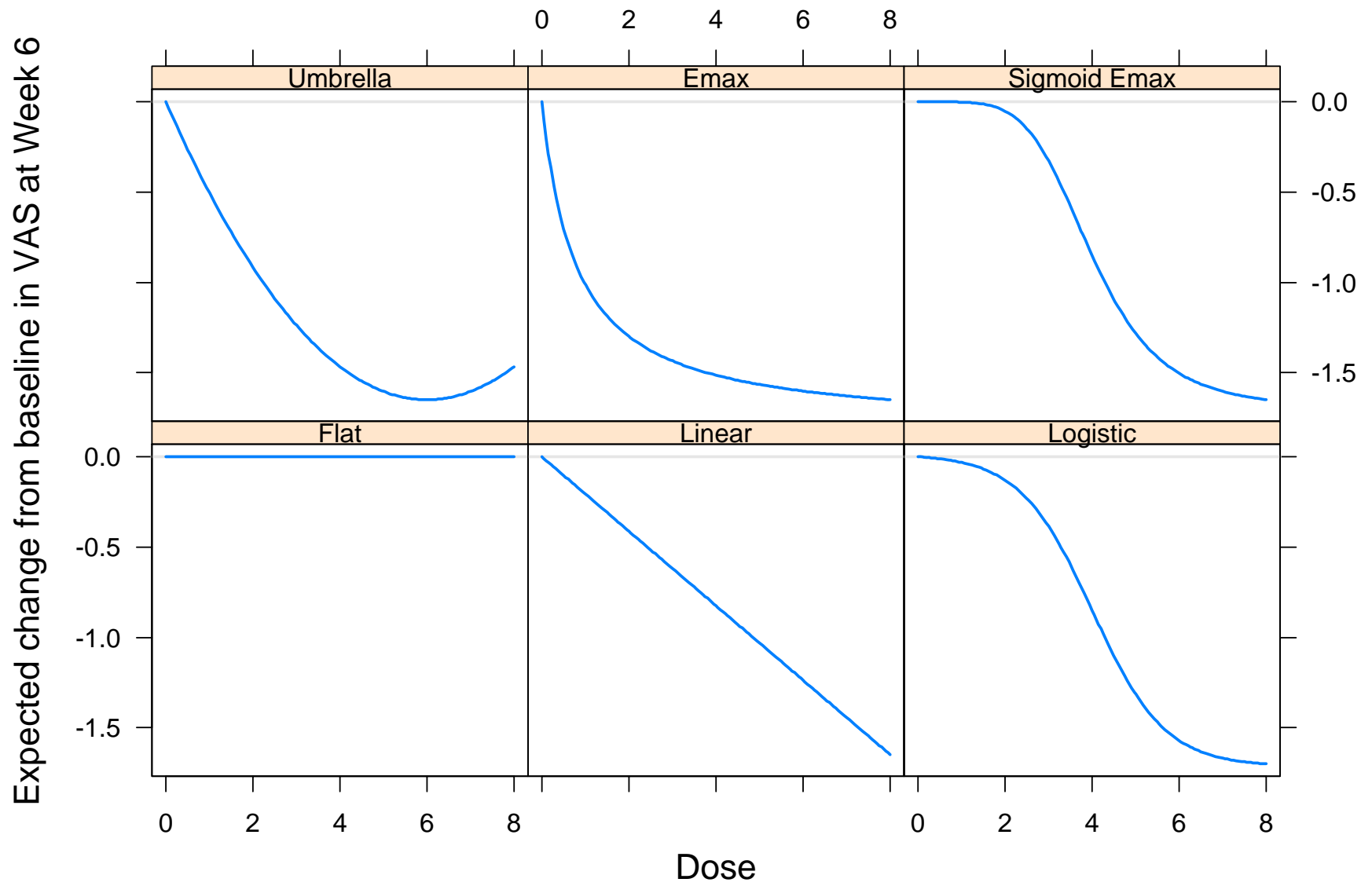
## Simulation study: design and assumptions

- Objective: proof-of-concept + dose finding for neuropathic pain
- Primary endpoint: change from baseline in pain score (VAS)
- Key questions:
  - is there evidence of a dose response
    - \* Significance level (one-sided FWER): 0.05
    - \* Clinically relevant change in VAS: 1.3
  - which dose(s) should be tested in large confirmatory trials
  - how well is the dose response (DR) curve estimated
- Study design scenarios:
  - Sample sizes: 150 and 250 patients
  - Number of doses: 5, 7, and 9 doses <sup>a</sup>

---

<sup>a</sup>5 doses (0, 2, 4, 6, 8), 7 doses (0, 2, 4, 6, 8), and 9 doses (0, 1, . . . , 8)

# Dose response profiles



## Measuring performance

---

- Probability of identifying dose response:  $Pr(DR)$
- Probability of identifying clinical relevance and selecting a dose for confirmatory phase:  $Pr(dose)$
- Dose selection: Distribution of selected doses (rounded to nearest integer, if continuous estimate possible)

## Measuring performance (contd.)

- Target dose interval – doses that produce effect within  $\pm 10\%$  of target effect  $\Delta$

Model	Target dose		Target interval	
	actual	rounded	actual	rounded
Linear	6.30	6	(5.67, 6.93)	{6,7}
Logistic	4.96	5	(4.65, 5.35)	{5}
Umbrella	3.24	3	(2.76, 3.81)	{3,4}
E <sub>max</sub>	2.00	2	(1.44, 2.95)	{2,3}
Sig-E <sub>max</sub>	5.06	5	(4.68, 5.58)	{5}

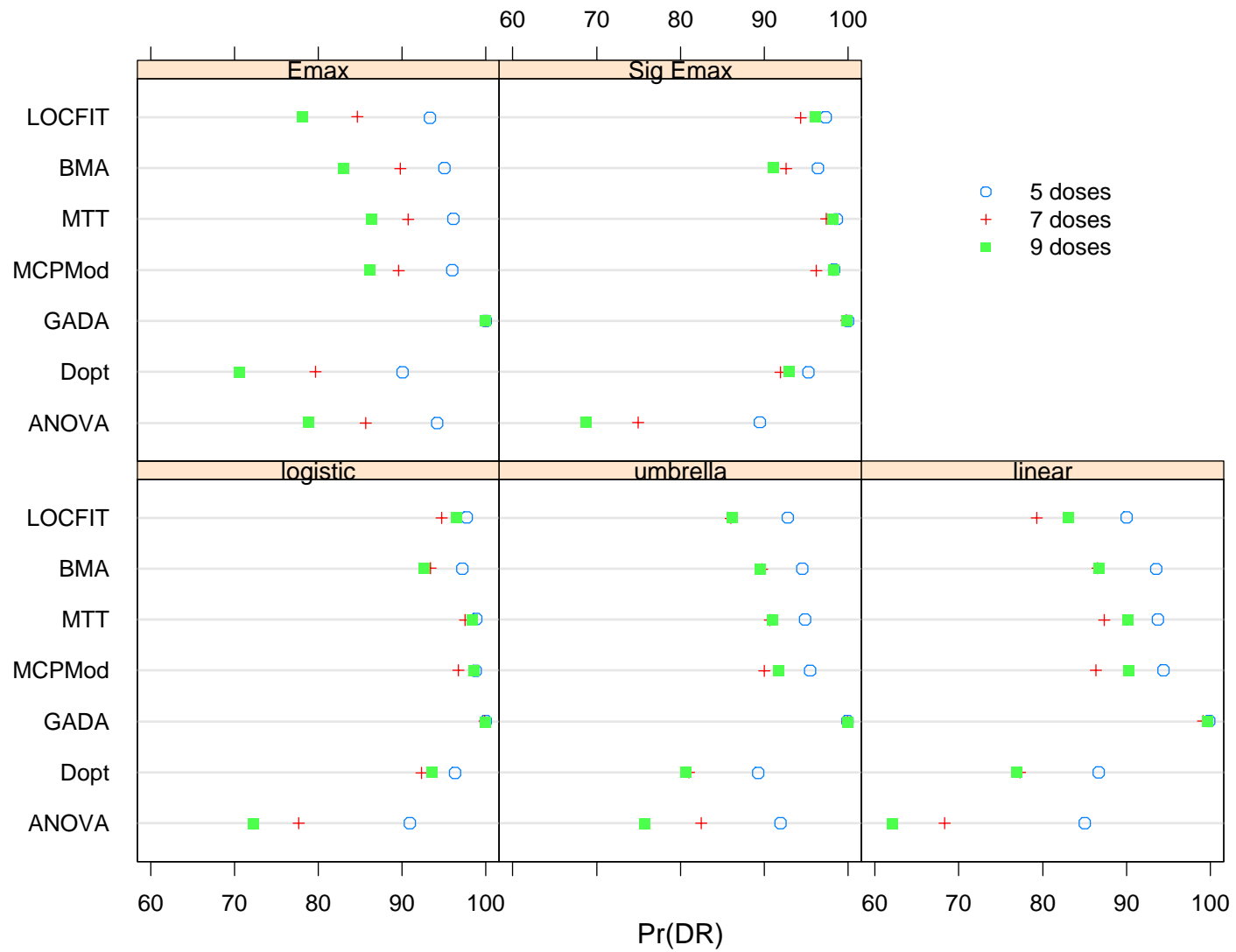
- Probabilities of under-, over-, and correct interval estimation:

$$P^- = P(\hat{d}_{\text{targ}} < d_{\text{min}}), \quad P^+ = P(\hat{d}_{\text{targ}} > d_{\text{min}}),$$

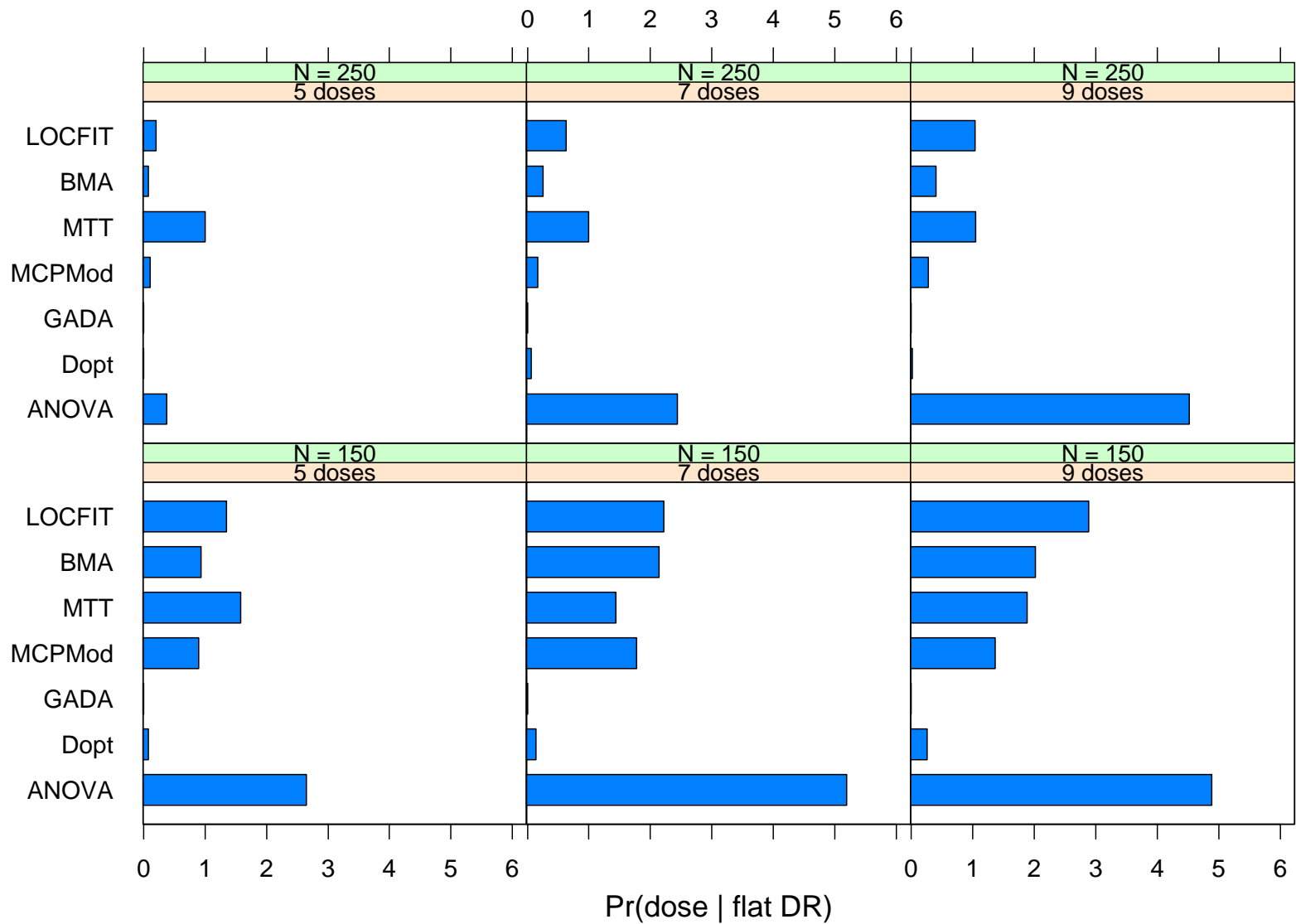
$$P^\circ = 1 - (P^- + P^+)$$

# Selected Simulation Results

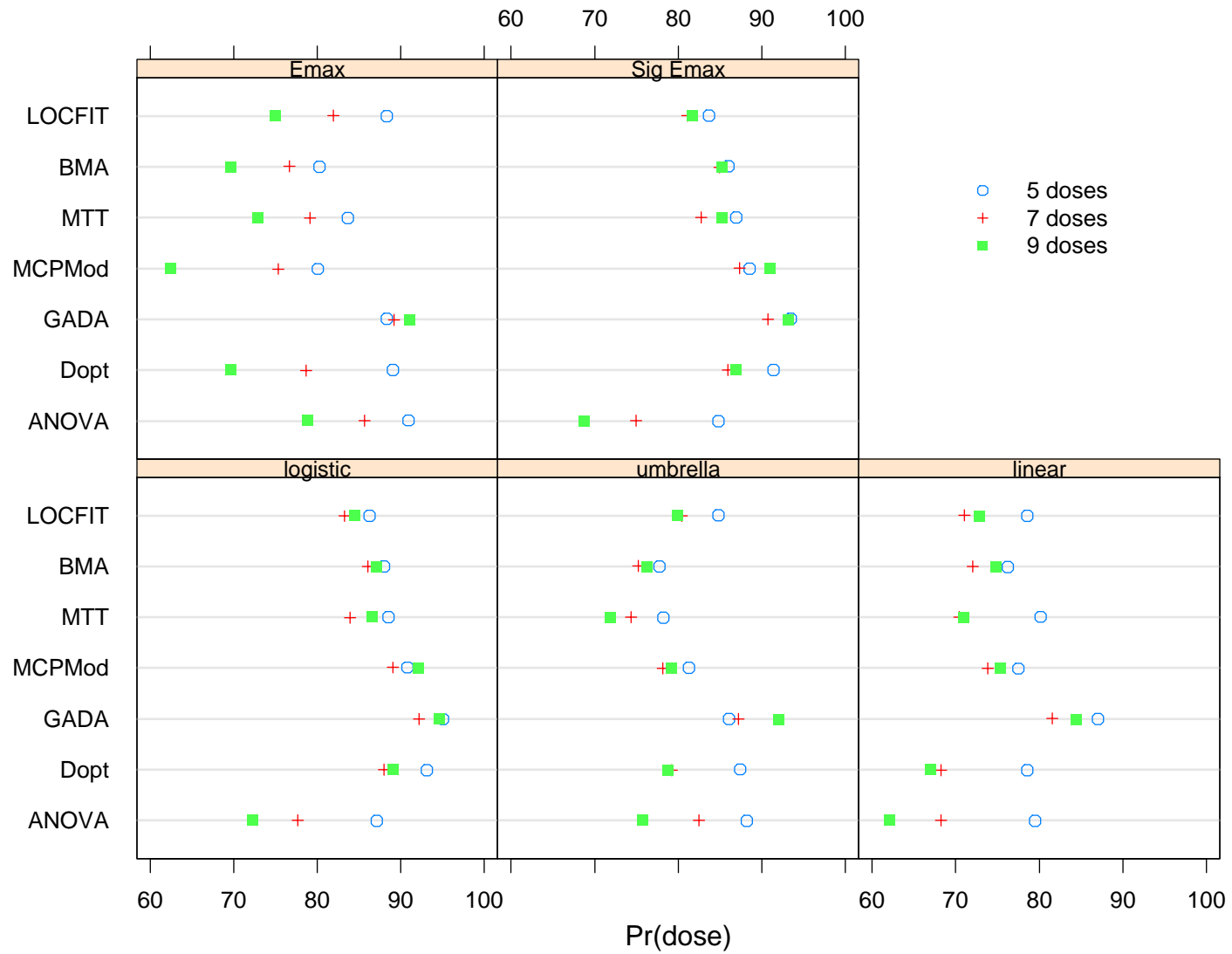
# Probability of identifying DR, N = 150



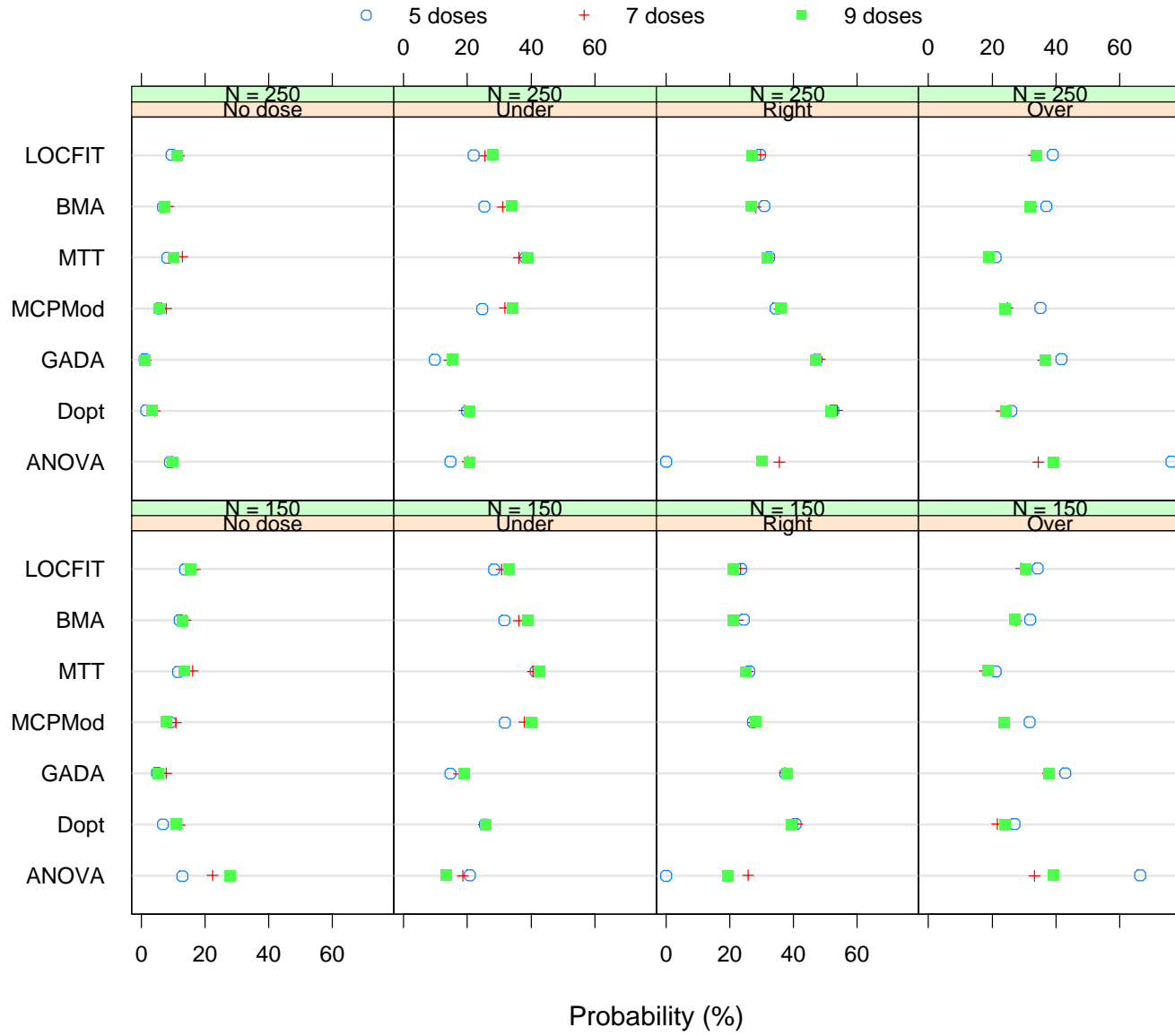
# Probability dose selection – flat DR



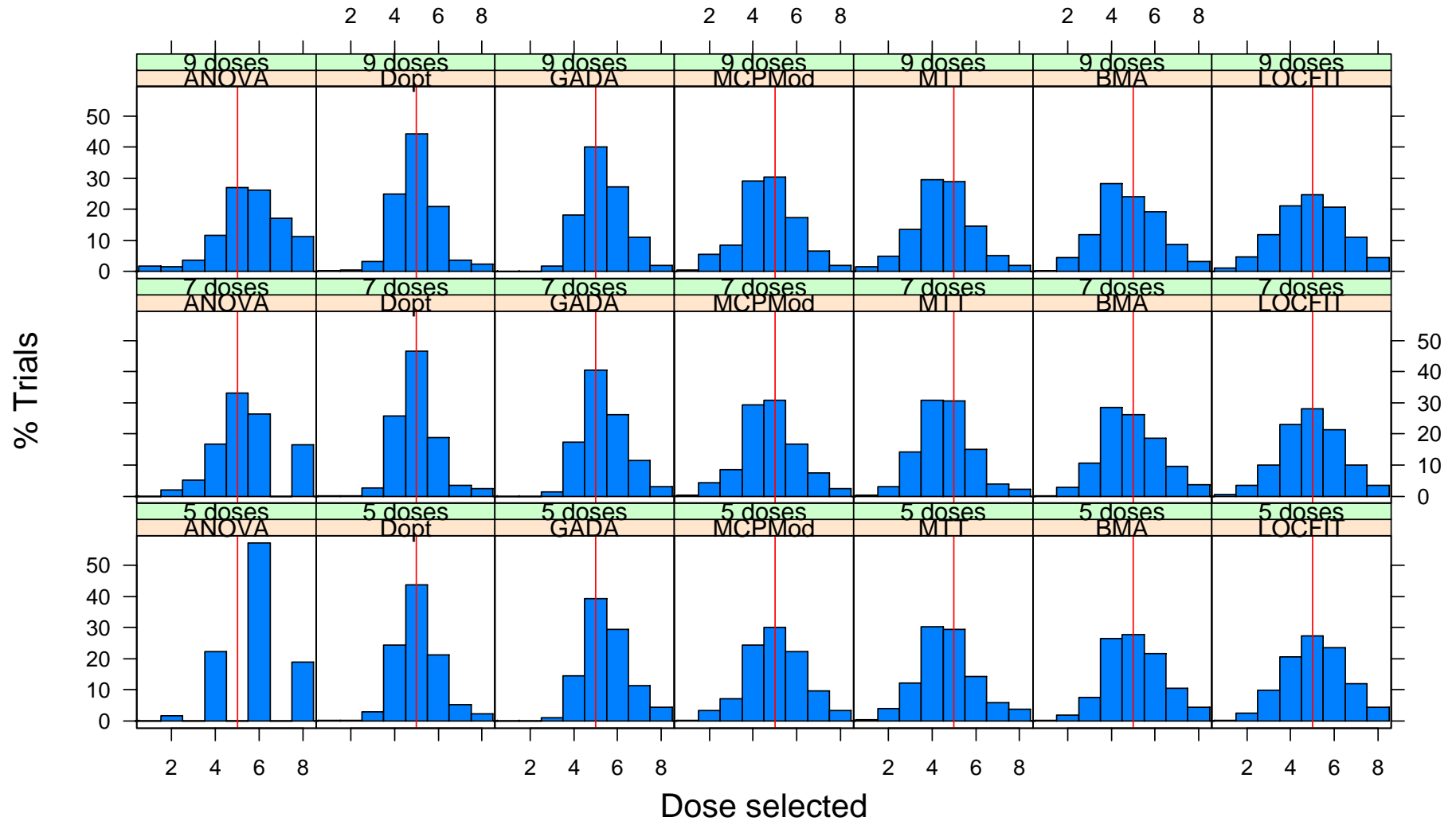
# Probability dose selection, N = 150



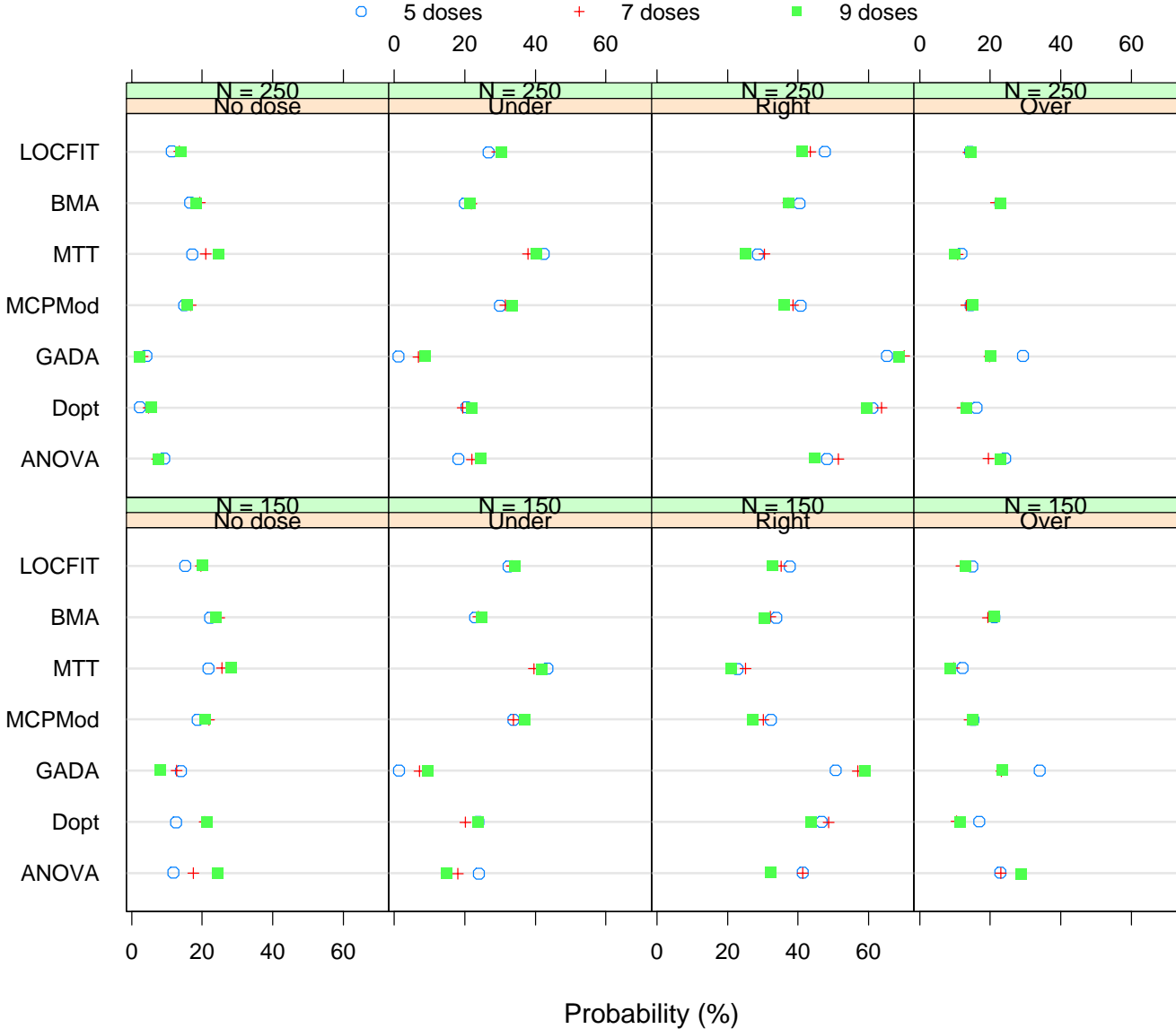
# Prob. of interval dose selection, Logistic model



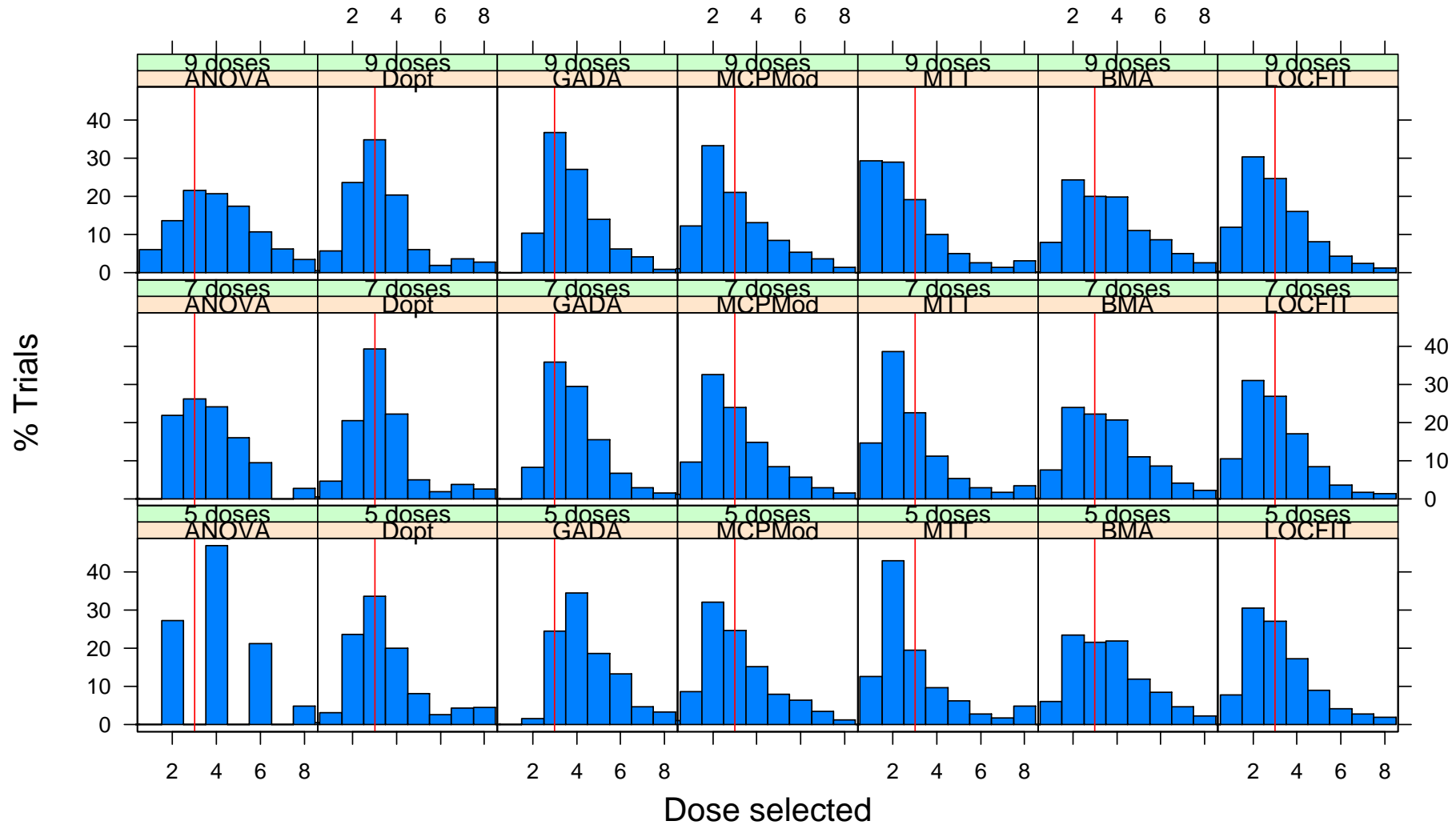
# Estimated dose distrib., Logistic model and N = 150



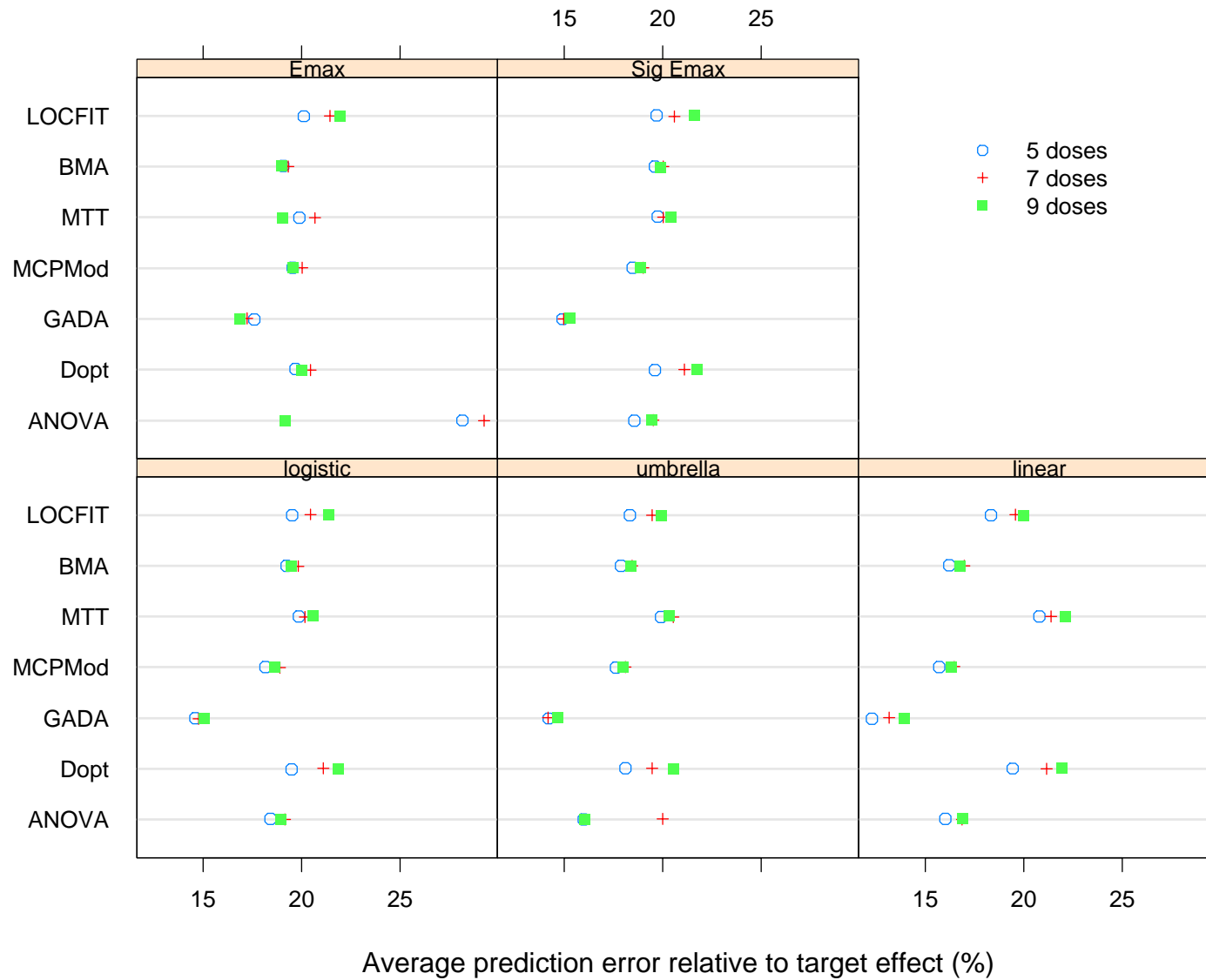
# Prob. of interval dose selection, Umbrella model



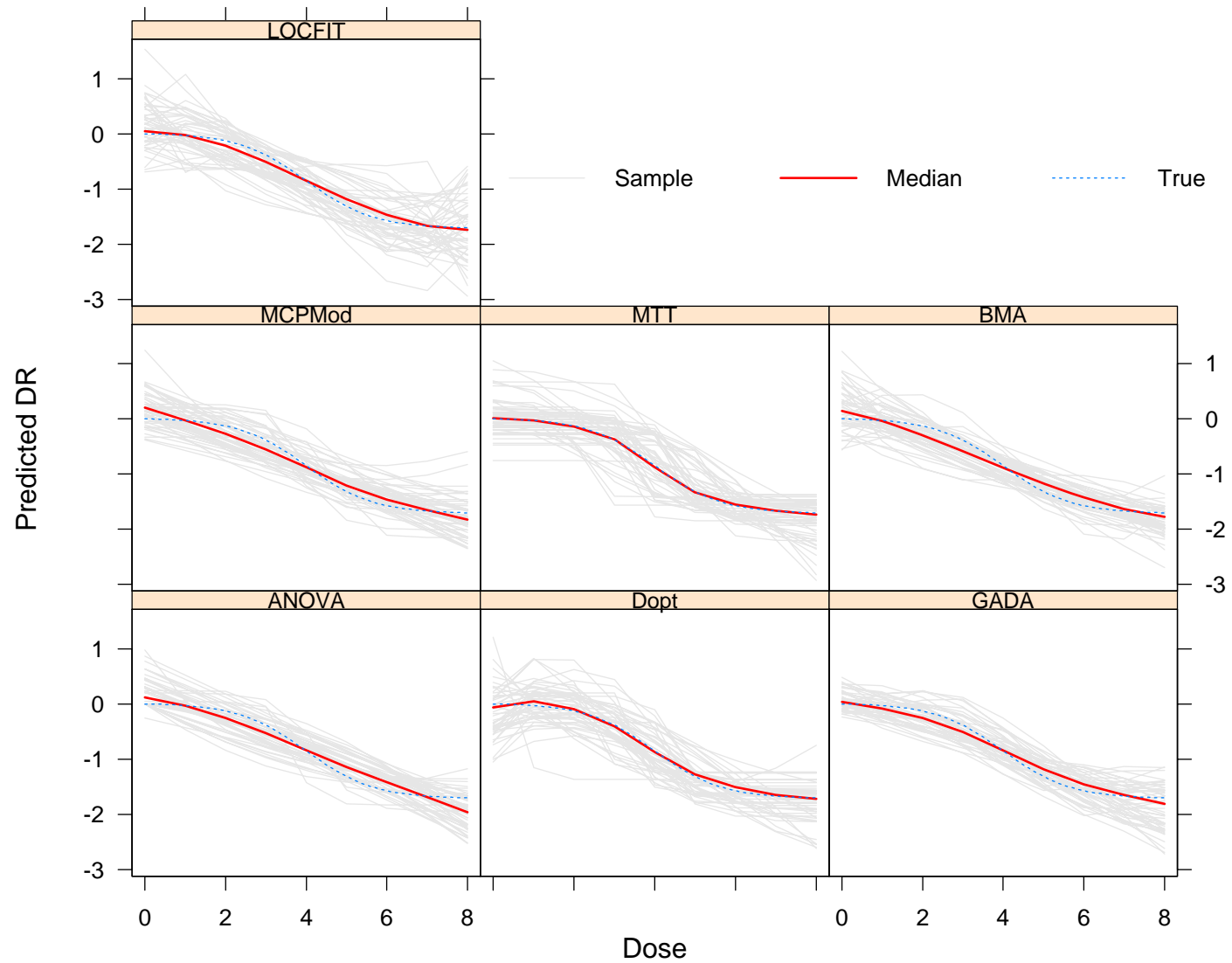
# Estimated dose distrib., Umbrella model and N = 150



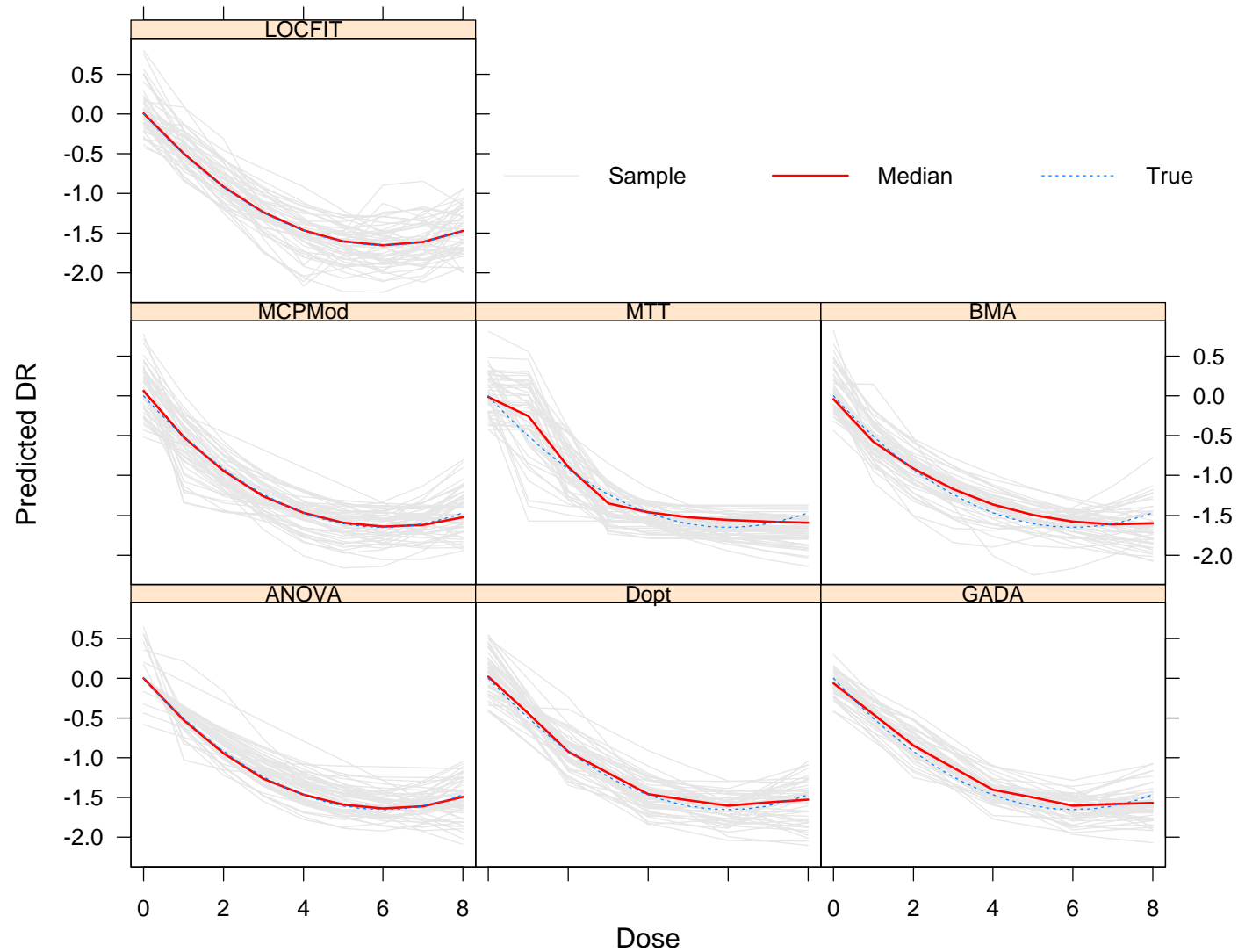
# Average prediction error per dose, N = 150



# Sample predicted curves: Logistic, 9 doses and N = 150



# Sample predicted curves: Umbrella, 5 doses and N = 250



## Conclusions

---

- Detecting DR is considerably easier than estimating it
- Current sample sizes for DF studies, based on power to detect DR, are inappropriate for dose selection and DR estimation
- None of methods had good performance in estimating dose in the correct target interval: maximum observed percentage of correct interval selection – 60%  $\implies$  larger  $N$  needed
- Adaptive dose-ranging methods (i.e., RDS) lead to gains in power to detect DR, precision to select target dose, and to estimate DR – greatest potential in the latter two
- GADA had best overall performance, especially on DR estimation

## Conclusions (contd.)

---

- Model-based methods have superior performance compared to methods based on hypothesis testing
- Number of doses larger than 5 does not seem to produce significant gains (provided overall  $N$  is fixed)  $\implies$  trade-off between more detail about DR and less precision at each dose
- In practice, need to balance gains associated with adaptive dose ranging designs approach against greater methodological and operational complexity

## Preliminary Recommendations

---

- Adaptive, model-based dose-ranging designs should be used routinely in drug development, as they can lead to substantial gains in performance over traditional DF methods
- Sample size calculations for Phase II studies should take into account desired precision of estimated target dose and possibly also estimated DR (current methods are not appropriate)
- When resulting sample size is not feasible, should consider selecting two or three doses for confirmatory phase to increase likelihood of including “correct” dose – adaptive designs could be used in confirmatory phase for greater efficiency (e.g., dropping less efficient doses earlier)

## Preliminary Recommendations (contd.)

---

- Proof-of-concept (PoC) and dose selection should be combined, when feasible, into one seamless trial
- Early stopping rules, for both efficacy and futility, should be used when feasible to allow greater efficiency in adaptive designs – Bayesian methods are particularly well-suited for this purpose
- Trial simulations should be used to determine appropriate sample sizes, as well as for estimating operational characteristics of designs/methods under consideration

# References

Bretz, F., Pinheiro, J. and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies, *Biometrics* **61**(3): 738–748.

Krams, M., Lees, K. R. and Berry, D. A. (2005). The past is the future: Innovative designs in acute stroke therapy trials, *Stroke* **36**(6): 1341–7.